# Trypanosoma brucei glycoproteins contain novel giant poly-N-acetyllactosamine carbohydrate chains

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 2, Issue of January 14, pp. 865–871, 2005
Printed in U.S.A.


# *Trypanosoma brucei* Glycoproteins Contain Novel Giant Poly-*N*-acetyllactosamine Carbohydrate Chains*

Received for publication, September 27, 2004, and in revised form, October 25, 2004
Published, JBC Papers in Press, October 27, 2004, DOI 10.1074/jbc.M411061200

**Abdelmadjid Atrih‡, Julia M. Richardson§, Alan R. Prescott‡, and Michael A. J. Ferguson‡¶**

*From the ‡Division of Biological Chemistry and Molecular Microbiology, the School of Life Sciences, University of Dundee, Dundee DD1 5EH, and the §School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, Scotland, United Kingdom*

**The flagellar pocket of the bloodstream form of the African sleeping sickness parasite *Trypanosoma brucei* contains material that binds the β-D-galactose-specific lectin ricin (Brickman, M. J., and Balber, A. E. (1990) *J. Protozool.* 37, 219–224). Glycoproteins were solubilized from bloodstream form *T. brucei* cells in 8 M urea and 3% SDS and purified by ricin affinity chromatography. Essentially all binding of ricin to these glycoproteins was abrogated by treatment with peptide *N*-glycosidase, showing that the ricin ligands are attached to glycoproteins via *N*-glycosidic linkages to asparagine residues. Glycans released by peptide *N*-glycosidase were resolved by Bio-Gel P-4 gel filtration into two fractions: a low molecular mass mannose-rich fraction and a high molecular mass galactose and *N*-acetylglucosamine-rich fraction. The latter fraction was further separated by high pH anion exchange chromatography and analyzed by gas chromatography mass spectrometry, one- and two-dimensional NMR, electrospray mass spectrometry, and methylation linkage analysis. The high molecular mass ricin-binding *N*-glycans are based on a conventional Manα1–3(Manα1–6)Manβ1–4GlcNAcβ1–4GlcNAc core structure and contain poly-*N*-acetyllactosamine chains. A significant proportion of these structures are extremely large and of unusual structure. They contain an average of 54 *N*-acetyllactosamine (Galβ1–4GlcNAc) repeats per glycan, linked mostly by -4GlcNAcβ1–6Galβ1- interrepeat linkages, with an average of one -4GlcNAcβ1–3(-4GlcNAcβ1–6)Galβ1- branch point in every six repeats. These structures, which also bind tomato lectin, are twice the size reported for the largest mammalian poly-*N*-acetyllactosamine *N*-linked glycans and also differ in their preponderance of -4GlcNAcβ1–6Galβ1- over -4GlcNacβ1–3Galβ1- interrepeat linkages. Molecular modeling suggests that -4GlcNAcβ1–6Galβ1- interrepeat linkages produce relatively compact structures that may give these giant *N*-linked glycans unique physicochemical properties. Fluorescence microscopy using fluorescein isothiocyanate-ricin indicates that ricin ligands are located mainly in the flagellar pocket and in the endosomal/lysosomal system of the trypanosome.**

African trypanosomes are tsetse fly-transmitted parasitic protozoa responsible for sleeping sickness in man and nagana in livestock. In the mammalian host, bloodstream form *Trypanosoma brucei* live and divide extracellularly in the blood, lymph, and interstitial fluids. A densely packed layer of variant surface glycoprotein (VSG)[1] covers the surface of the parasites (1). Each trypanosome contains several hundred VSG genes that encode immunologically distinct molecules, and the host is unable to clear the infection because of antigenic variation (2, 3). The flagellar pocket, a deep invagination of the plasma membrane, is the only known site for endocytosis and the secretion of proteins from the parasite, and this organelle is also involved in the recycling/sorting of proteins to and from the cell surface (4–11). The fractional volume of the flagellar pocket in the bloodstream form parasite is ~3 times larger than that in the insect-dwelling procyclic form trypanosome (12). Large coated endocytic vesicles are present only in the bloodstream form of the parasite (5, 13–15). This suggests that some components in the flagellar pocket are stage-specific and that this organelle plays a crucial role in general cellular function and pathogenesis (9, 13, 16, 17). Thus, flagellar pocket glycoproteins represent an important biological interface between host and parasite.

Despite the pivotal physiological role of the flagellar pocket in bloodstream form *T. brucei*, very little is known about its biochemical constituents. Although a small number of glycoproteins that reside in the flagellar pocket and lysosomal/endosomal system have been identified, very little is known about the fine chemical structures of their glycan moieties (16, 19–21). One known feature is the presence of electron-dense material in the flagellar pocket that binds ricin (18) and tomato lectin (19). The binding of tomato lectin suggests the presence of linear poly-*N*-acetyllactosamine (poly-LacNAc) oligosaccharides (19). In mammalian systems, poly-LacNAc-containing glycans that bind tomato lectin occur both on cell surface and lysosomal membrane glycoproteins (22–24) and consist of up to about 26 Galβ1–4GlcNAc units linked together principally by -4GlcNAcβ1–3Galβ1- linkages, occasional -4GlcNAcβ1–6Gal linkages, and some -4GlcNAcβ1–6(-4GlcNAcβ1–3)Galβ1- branch points. Medium sized poly-LacNAc chains of this type (with up to 8 repeats) have been identified on certain type 2 VSGs of *T. brucei* (25).

* This work was supported by Wellcome Trust Grants 62387 and 71463. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: University of Dundee School of Life Sciences, Wellcome Trust Biocentre, Dow St., Dundee DD1 5EH, Scotland, United Kingdom. E-mail: M.a.j.ferguson@dundee.ac.uk.

[1] The abbreviations used are: VSG, variant surface glycoprotein; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; ES-MS, electrospray ionization-mass spectrometry; ES-MS/MS, electrospray ionization-tandem mass spectrometry; FITC, fluorescein isothiocyanate; GC-MS, gas chromatography-mass spectrometry; HPAEC, high pH anion exchange chromatography; LacNAc, *N*-acetyllactosamine (Galβ1–4GlcNAc); MALDI-MS, matrix-assisted laser desorption/ionization-mass spectrometry; PBS, phosphate-buffered saline; PMAA, partially methylated alditol acetate; PNGase-F, peptide *N*-glycosidase F; poly-LacNAc, poly-*N*-acetyllactosamine; ROESY, rotating frame nuclear Overhauser spectroscopy; TOCSY, total correlation spectroscopy; VLPL, very large poly-*N*-acetyllactosamine.

To begin to understand the physiological role of the flagellar pocket in bloodstream form *T. brucei,* we decided to examine the ricin-binding glycoproteins in this organism. We report a procedure for extracting glycoproteins using ricin coupled to agarose and describe the unusual size and structure of poly-LacNAc glycans released by peptide *N*-glycosidase-F (PNGase-F). We also show the localization of ricin ligands in the trypanosome. The unique sizes and unusual structure of the poly-LacNAc chains are discussed.

## MATERIALS AND METHODS

*Extraction of Ricin-binding Glycoproteins, SDS-PAGE, and Western Blotting*—Bloodstream form *T. brucei* (strain 427, variant 117), were isolated from infected rats and purified over DEAE-cellulose (26). Parasites were subjected to hypotonic lysis to release cytosolic components as well as the majority of the VSG coat as soluble form VSG (26). The cell ghost pellets ($10^{11}$ cell equivalents) were solubilized in 50 ml of 8 M urea, 3% SDS, 50 mM Tris-HCl, pH 6.8. The SDS/urea extract was diluted 50 times in buffer A (50 mM Tris-HCl, pH 6.8, 400 mM NaCl, and 0.8% Triton X-100, 0.1 mM 1-chloro-3-tosylamido-7-amino-2-heptanone, 1 $\mu$g/ml leupeptin, and 0.1% sodium azide) to allow affinity chromatography on ricin-agarose (Vector Laboratories). Ricin-coupled agarose (4-ml packed volume) was added to the suspension and rotated gently overnight at 4 °C. Ricin-agarose was recovered by gentle centrifugation and packed into a 15 × 1-cm column. The column was washed with 5 volumes of buffer A, and bound material was eluted with 30 mg/ml lactose and 30 mg/ml Gal in 4-fold diluted buffer A. Aliquots of the eluted fractions were separated on NuPAGE bis-Tris 4–12% gradient gels (Invitrogen). After electrophoresis, glycoproteins were transferred onto nitrocellulose membrane (Hybond, Amersham Biosciences) at 45 mA for 1 h in a Hoefer SemiPhor semidry transfer unit. After blocking for 1 h in buffer B (Tris-buffered saline, pH 7.4, containing 0.05% Nonidet P-40, and 0.25% bovine serum albumin) at room temperature, membranes were incubated with horseradish peroxidase-conjugated ricin (diluted 1:10,000 in buffer B) for 1 h. The membrane was washed three times in buffer B and twice with PBS. Visualization was by enhanced chemiluminescence (Amersham Biosciences). Fractions containing ricin-binding glycoproteins were pooled, concentrated 5-fold using a centrifugal concentrator, and diafiltered against water to remove most of the lactose and Gal. The concentrated extract was stored at −20 °C before use. An aliquot of this concentrated extract was also separated on a NuPAGE bis-Tris 4–12% gradient gel and stained for carbohydrate by periodic acid-Schiff. After electrophoresis, the gel was rinsed with water and incubated with fixing solution (acetic acid/methanol/water; 10:35:25) for 20 min. The gel was washed three times (5 min each) with water and incubated with oxidation solution (1% sodium periodic acid in 3% acetic acid) for 30 min. The gel was washed three times (5 min each) with water and incubated for 1 h with Schiff's reagent (Sigma). Finally, the gel was incubated for 30 min with a reducing solution of 1% sodium metabisulfite and washed six times with water.

*N-Glycan Release and Purification*—To release *N*-glycans from proteins, the ricin-binding glycoproteins from $1 \times 10^{11}$ cells in ~1.5 ml were treated with 7,000 units of PNGase-F (New England BioLabs) overnight at 37 °C. The PNGase-F-treated fraction was made 50 mM with respect to potassium acetate and left overnight at 4 °C to precipitate SDS from the buffer. Triton X-100 was exhaustively extracted from the supernatant with toluene (typically 15–20 extractions with 3 volumes of toluene). Traces of solvent were removed by evaporation with nitrogen, and precipitated deglycosylated proteins at the water/toluene interface were discarded by centrifugation. *N*-Glycans were freeze dried, redissolved in water, and applied to a Bio-Gel P-4 gel filtration column (40 × 1 cm) eluted with water at 6 ml/h. Aliquots (10 $\mu$l) of 0.5-ml fractions were mixed with 200 pmol of *scyllo*-inositol and subjected to methanolysis, trimethylsilylation, and GC-MS monosaccharide composition analysis (27). Fractions that eluted in the void volume of the column, which contained mainly Gal and GlcNAc, were pooled and desalted using a PD-10 column (Amersham Biosciences) eluted with Milli-Q water. This fraction, referred to as the "total poly-LacNAc fraction," was further fractionated by high pH anion exchange chromatography (HPAEC) using a Dionex CarboPac PA-100 column (4 × 250 mm). The column was equilibrated with 95% buffer A (100 mM NaOH) and 5% buffer B (380 mM sodium acetate in 100 mM NaOH) for 20 min at a flow rate of 0.6 ml/min. *N*-Glycans were separated using a linear gradient of 5–40% buffer B over 90 min at 0.6 ml/min. *N*-Glycans were detected with a pulse-amperiometric detector, and sodium ions were removed from the eluate using an on-line Dionex ARRS unit (27). Glycans were collected individually or in fractions at the detector outlet and desalted by passage through a column of 0.5 ml of Dowex AG 50-$H^+$ over 0.5 ml of Dowex AG 3-$OH^-$ and elution with 4 ml of water. The eluates were freeze-dried and redissolved in water. The unresolved fraction at the end of the Dionex HPAEC gradient (see Fig. 3) is referred to as the "very large poly-LacNAc" (VLPL) fraction.

*Mass Spectrometric Analysis of N-Glycans*—Aliquots (1 $\mu$l) of glycans resolved on Dionex HPAEC, desalted and redissolved in Milli-Q water, were mixed with 1 $\mu$l of acetonitrile containing 2% formic acid and loaded into nanospray tips (Micromass type F) for ES-MS and ES-MS/MS. Samples were analyzed in positive ion mode with capillary and cone voltages of 0.9 kV and 30 V, respectively, using a Micromass Quattro Ultima triple quadrupole mass spectrometer and/or a Micromass Q-Tof2 orthogonal quadrupole-time of flight mass spectrometer (Micromass, Manchester, UK). All spectra were collected and processed with MassLynx software.

*Methylation Linkage Analysis*—Methylation linkage analysis was carried out on the VLPL fraction that eluted at the end of Dionex HPAEC gradient. Approximately 5 nmol of material was used for each methylation linkage analysis. Glycans were converted to constituent monosaccharides in the form of partially methylated alditol acetates (PMAAs) and analyzed by GC-MS, as described previously (27). To estimate the stoichiometry of the PMAAs, authentic standards (*i.e.* NA2 asialo-bi-biantennary *N*-linked glycan, lacto-*N*-neohexaose, galactotriose (Dextra) and Man$\alpha$1–2Man$\alpha$1–6(Gal$\alpha$1–6Gal$\alpha$1–3)Man$\alpha$1–4AHM (28)) with glycosidic bonds present in the VLPL fraction (*i.e.* 4-*O*-substituted-GlcNAc, terminal-Gal, 3-*O*-substituted-Gal, 6-*O*-substituted-Gal, 3,6-di-*O*-substituted-Gal, 2-*O*-substituted Man, 3,6-di-*O*-substituted-Man) were analyzed simultaneously to determine total ion current molar relative response factors for the different residue types.

*NMR Analysis of the VLPL Fraction*—The VLPL fraction purified by Dionex HPAEC was dissolved in $^2H_2O$ and analyzed by one- and two-dimensional $^1H$ NMR. All NMR spectra were acquired at a probe temperature of 298 K on a Bruker Avance spectrometer operating at 600 MHz for $^1H$. To assign the proton resonances, two-dimensional homonuclear correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY), with a mixing time of 120 ms, were performed. In addition, rotating frame nuclear Overhauser spectroscopy (ROESY), with a mixing time of 300 ms, was carried out to determine the connections between carbohydrate residues. An estimate of the proportion of LacNAc repeats to core residues was obtained by integration of the anomeric protons in the one-dimensional NMR spectrum using the program XWIN-NMR (Bruker).

*Dynamic Light Scattering Analysis of the VLPL Fraction*—The VLPL fraction was analyzed on a Proterion Corporation model 99E DynaPro instrument with MS800 optics fitted with a flow cell. Approximately 300 $\mu$l of 0.2-$\mu$m filtered sample was injected into the flow cell, and 20 acquisitions were recorded using the model for branched polysaccharides. Each acquisition was the average of five 1-s readings.

*Tomato Lectin Chromatography*—Approximately 18 nmol of the total poly-LacNAc fraction that eluted in the void volume of the Bio-Gel P-4 column (see Fig. 2) were mixed with 4 ml (bed volume) of tomato lectin-agarose (Vector). The agarose was equilibrated with buffer A (25 mM Tris-HCl pH 7.0, 150 mM NaCl, 0.1 mM $CaCl_2$, and 0.1% sodium azide), and the mixture was left to rotate gently overnight at 4 °C. The agarose was washed by centrifugation with 16 ml of equilibration buffer, and bound oligosaccharides were eluted using a mixture of chitooligosaccharides (20 mg/ml of tri-*N*-acetylchitotriose and tetra--*N*-acetylchitotetraose in buffer A). Glycans present in the flow-through, wash, and elution fractions were pooled separately and purified using a Bio-Gel P-4 gel filtration column. Void volume fractions containing glycans were then desalted using a PD-10 column (Amersham Biosciences), dried, and redissolved in Milli-Q water. The glycan content of the three different fractions was examined by GC-MS. Glycans present in the flow-through and wash fractions were examined by ES-MS to determine the nature of the structures that did not bind to tomato lectin. The glycans that were eluted from the tomato lectin-agarose were analyzed by Dionex HPAEC, as described above.

*Fluorescence Microscopy*—Bloodstream form *T. brucei* (strain 427, variant 117) were isolated from infected mice, purified over DEAE-cellulose, washed, and resuspended in ice-cold trypanosome dilution buffer (25 mM KCl, 400 mM NaCl, 5 mM $MgSO_4$, 100 mM $Na_2HPO_4$, 10 mM $NaH_2PO_4$, and 100 mM glucose) at a final concentration of $2 \times 10^7$ cells/ml. The parasites were air-dried on coverslips (13 mm) at room temperature, fixed in 4% paraformaldehyde for 30 min at 4 °C, and then washed four times with PBS (Sigma). For visualization of ricin binding

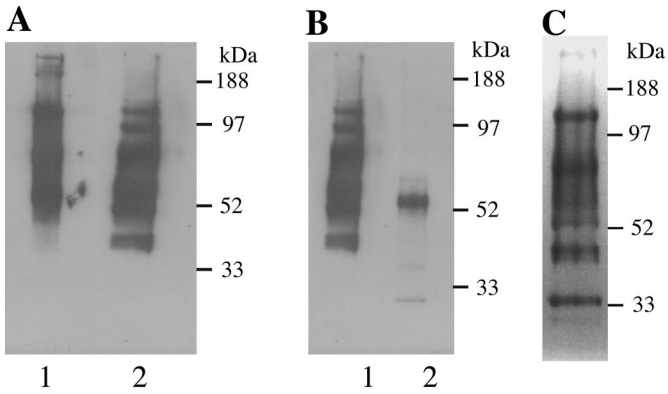


FIG. 1. **Ricin blot analysis of glycoproteins from bloodstream form *T. brucei*.** Glycoproteins were eluted from a ricin-agarose column using a mixture of galactose and lactose. After electrophoresis, glycoproteins were either transferred onto nitrocellulose membrane and probed with horseradish peroxidase-conjugated ricin (*A* and B) or stained for carbohydrate with periodic acid-Schiff (*C*). *A*: *lane 1*, native extract; *lane 2*, extract reduced with dithiothreitol. *B*: *lane 1*, reduced extract; *lane 2*, reduced extract after digestion with PNGase-F. *C*: reduced extract. The positions of molecular mass standards are indicated.

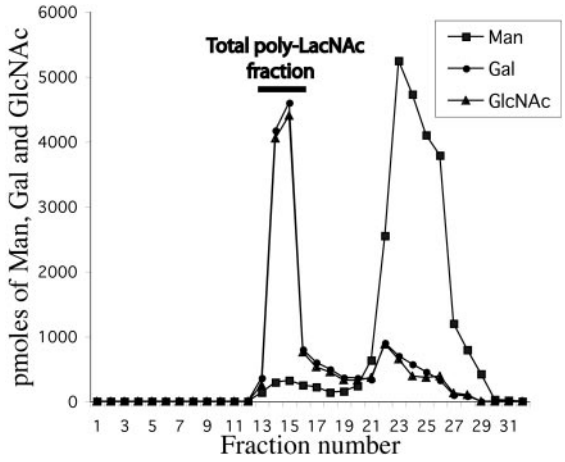


FIG. 2. **Bio-Gel P-4 chromatogram of *N*-glycans obtained from ricin-binding glycoproteins.** Glycans released from ricin-binding glycoproteins by the action of PNGase-F were purified from detergents and salts, as described under "Materials and Methods," and separated on a Bio-Gel P-4 column. The monosaccharide compositions of the fractions were determined by GC-MS. The void volume total poly-LacNAc fraction is indicated with a *bar*.

sites, fixed parasites were treated with 0.1% bovine serum albumin in PBS for 1 h to block nonspecific binding. Cells were stained with FITC/ricin (1/500) in PBS (with and without 30 mg/ml lactose and galactose) for 1 h and then washed six times with PBS. Specimens were mounted in Hydromount, and single optical sections were collected on a Zeiss 510 META confocal microscope (alpha-Plan-Fluor ×100).

*Molecular Modeling*—A model of -4GlcNAc$\beta$1-3Gal$\beta$1-linked poly-LacNAc containing five LacNAc repeats was built in Insight (Accelrys). A series of 10 random geometries was then computed, and each was subjected to dynamic simulated annealing using Discover (Accelrys). Similarly, 10 random geometry models of -4GlcNAc$\beta$1-6Gal$\beta$1-linked poly-LacNAc, also containing five LacNAc repeats, were generated and subjected to simulated annealing.

## RESULTS

*Extraction of Ricin-binding Glycoproteins*—To extract all types of membrane-associated glycoconjugates, osmotically lysed *T. brucei* cells were completely solubilized in 8 M urea and 3% SDS for 30 min before dilution of the extract and addition of ricin coupled to agarose. Rechromatography of the unbound fraction showed that the majority of the ricin-binding material was captured in the first round of chromatography (data not shown). The native glycoproteins eluted from the ricin-coupled agarose range in apparent molecular mass from 35 to more than 188 kDa (Fig. 1*A*, *lane 1*). Reduction with 1 M dithiothreitol resulted in the disappearance of bands over 188 kDa and the appearance of smaller molecular mass glycoproteins (Fig. 1*A*, *lane 2*), suggesting disulfide linkages between some of the constituent glycoproteins. PNGase-F digestion of the ricin-binding glycoproteins revealed that almost all of the ricin-binding glycans are *N*-linked structures (Fig. 1*B*). The remaining ricin binding band at about 50 kDa is most likely residual VSG, a proportion of which is known to express ricin-binding terminal $\beta$-galactosidase as a side chain on its glycosylphosphatidylinositol anchor (28, 29). Treatment of the PNGase-F-treated sample with ice-cold aqueous HF, which removes glycosylphosphatidylinositols from proteins (27), abrogated ricin binding to this band (data not shown). The reduced sample was also analyzed by SDS-PAGE and periodic acid-Schiff staining for carbohydrate (Fig. 1*C*). The pattern of staining is similar to that of the ricin Western blot (Fig. 1*A*, *lane 2* and Fig. 1*B*, *lane 1*) except that the band running below the 97 kDa molecular mass marker in the ricin blot is less pronounced in the periodic acid-Schiff stain, and the band at 35 kDa in the periodic acid-Schiff stain is not apparent in the ricin blot. Presumably, the former has a relatively high terminal $\beta$-galactose to total carbohydrate ratio, and the latter does not contain terminal $\beta$-galactose itself but copurifies with other ricin-binding glycoproteins.

Glycoproteins from the supernatant of the osmotically lysed parasites were also purified using agarose-coupled ricin. However, this fraction contained mostly soluble form VSG, which is known to be released from the cell surface during osmotic lysis by enzymatic cleavage of the glycosylphosphatidylinositol anchor (26, 28) and was not analyzed further.

*Purification of VLPL Glycans*—*N*-Linked glycans released from the ricin-binding glycoproteins by PNGase-F were separated by Bio-Gel P-4 gel filtration chromatography. The monosaccharide compositions of fractions collected from the Bio-Gel P-4 column were analyzed by GC-MS. The Man, Gal, and GlcNAc contents were plotted against fraction number (Fig. 2). The *N*-glycans were resolved into two main fractions: a Gal/GlcNAc-(poly-*N*-LacNAc-) rich high molecular mass fraction eluting at the column void volume (the total poly-LacNAc fraction) and a Man-rich low molecular mass fraction eluting as a broad peak. This column also allowed the removal of the remaining Gal and lactose used in elution from ricin-coupled agarose.

A representative chromatogram of the total poly-LacNAc fraction separated by Dionex HPAEC is shown in Fig. 3*A*. Based on monosaccharide composition analysis, only the peaks eluting after 10 min contained carbohydrate. The poly-LacNAc fraction resolved into more than 20 discrete peaks. However, almost 70% of *N*-glycans eluted at the end of the gradient as large unresolved peak. This material was pooled as the VLPL fraction.

*Structural Analysis of the VLPL Fraction*—The VLPL fraction eluted from the Dionex HPAEC column was first subjected to monosaccharide compositional analysis by GC-MS. This fraction contains exclusively Man, Gal, and GlcNAc in the ratio 3:59:60. Methylation linkage analysis (Table I) revealed 2-*O*-substituted-Man and 3,6-disubstituted-Man and no other types of Man residue. Furthermore, one- and two-dimensional $^1$H NMR spectra of the same fraction after exchange into $^2H_2O$ (Fig. 4, *A* and *B*, Table II) revealed resonance values for $\alpha$Man, $\beta$Man and $\beta$GlcNAc residues (indicated with an *asterisk* in Table II) similar to those reported for the conserved $Man_3GlcNAc_2$ core of many *N*-linked glycans (30). Taken together, these data suggest that all or most of the structures present are based on a conventional biantennary core structure

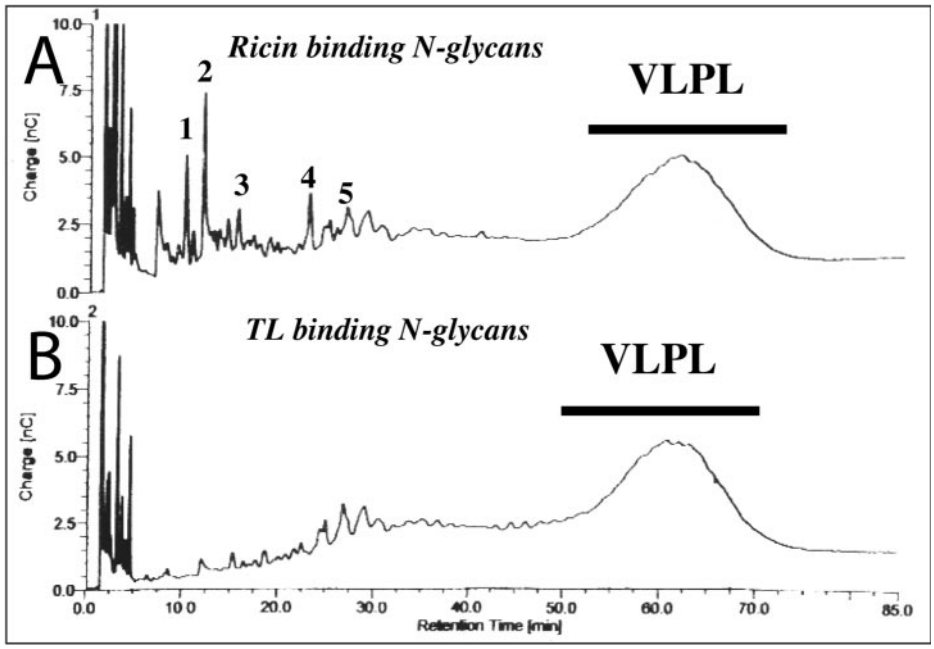


FIG. 3. **Dionex HPAEC separation of total poly-LacNAc fractions.** *A*, HPAEC chromatogram of glycans, detected by pulsed-amperiometric detection, of the total poly-LacNAc fraction isolated after Bio-Gel P-4 gel filtration (see Fig. 2). *B*, HPAEC chromatogram of the glycans of the total poly-LacNAc fraction which bind to tomato lectin-agarose. The VLPL is indicated with a *bar*.

TABLE I

*Quantitative GC-MS methylation linkage analysis of the VLPL fraction*

The VLPL fraction was methylated, hydrolyzed, reduced, and acetylated to yield PMAA derivatives for analysis by GC-MS. Derivatives were identified by retention time and electron impact mass spectra and quantified by integration of the total ion current using molar response factors deduced from concurrent analysis of authentic standards (see "Materials and Methods").

| PMAA derivative | Origin | Amount[a] |
|---|---|---|
| 2,4-Di-*O*-methyl-1,3,5,6-tetra-*O*-acetyl-1-[$^2$H]mannitol | 3,6-Di-*O*-substituted Man | 1 |
| 3,4,6-Tri-*O*-methyl-1,2,5-tri-*O*-acetyl-1-[$^2$H]mannitol | 2-*O*-Substituted Man | 2 |
| 2,3,4,6-Tetra-*O*-methyl-1,5-di-*O*-acetyl-1-[$^2$H]galactitol | Terminal Gal | 12 |
| 2,4,6-Tri-*O*-methyl-1,3,5-tri-*O*-acetyl-1-[$^2$H]galactitol | 3-*O*-Substituted Gal | 3 |
| 2,3,4-Tri-*O*-methyl-1,5,6-tri-*O*-acetyl-1-[$^2$H]galactitol | 6-O-Substituted Gal | 30 |
| 2,4-Di-*O*-methyl-1,3,5,6-tetra-*O*-acetyl-1-[$^2$H]galactitol | 3,6-Di-*O*-substituted Gal | 9 |
| 3,6-Di-*O*-methyl-1,4,5-tri-O-acetyl-2-*N*-methylacetamidoglucosaminitol | 4-*O*-substituted GlcNAc | 64 |

[a] Molar quantities relative to 3,6-di-*O*-substituted Man (one/glycan). The mean LacNAc repeat number/glycan estimated from the total Gal derivative content is 54.

of: R-2Man$\alpha$1–6(R′-2Man$\alpha$1–3)Man$\beta$1–4GlcNAc$\beta$1–4GlcNAc. The methylation linkage analysis and $^1$H NMR data are also in good agreement about the remaining Gal and GlcNAc residues. Both sets of data reveal the presence of large amounts of 4-*O*-substituted-($\beta$)GlcNAc and 6-*O*-substituted-($\beta$)Gal together with some 3,6-di-*O*-substituted-($\beta$)Gal and small amounts of 3-*O*-substituted-($\beta$)Gal. Strong connectivities in the ROESY spectrum show that most of the GlcNAc and Gal residues are present as Gal$\beta$1–4GlcNAc (LacNAc) repeats. Assuming 3 mol of Man/mol of glycan, we can estimate the amount of VLPL structures from the GC-MS monosaccharide composition analysis of this fraction. Thus, the measured yield of 60 nmol of Man/$10^{11}$ parasites suggests 20 nmol of VLPL glycans (equivalent to ~0.4 mg of carbohydrate)/$10^{11}$ parasites and 120,000 copies of VLPL glycan/parasite. These are minimum figures because we have not rigorously measured recovery at every stage of purification.

The only discrepancy between the composition, methylation linkage analysis and NMR data relates to stoichiometry. The compositional and methylation linkage data suggested an average of 59 and 54 LacNAc repeats, respectively, assuming 3 Man residues/glycan. However, integration of NMR resonances suggested ~22 repeats/glycan (Table II). Compositional and methylation analyses of the soluble material recovered from the NMR experiment agreed with the latter figure. However, we noticed that after vortexing the average repeat number (according to composition analysis) increased to 45, suggesting that some of the larger glycans had not been in free solution. We also realized that the process of rotary evaporation, used for exchange into $^2H_2O$, had also resulted in the selective loss of the larger oligosaccharides in the sample. Indeed, prolonged sonication of the vessel used to rotary evaporate the sample for NMR analysis with water recovered the larger oligosaccharides, as judged by compositional analysis and Dionex HPAEC (data not shown). We suggest that, like glycosaminoglycans, the larger oligosaccharides may have gel-like properties making their dissolution in water from a dried glass-like state particularly slow and dependent on sonication and giving them the tendency to come out of solution. To assess this point further, another VLPL preparation was exchanged into $^2H_2O$ but this time using freeze-drying. NMR analysis showed qualitatively exactly the same resonances as before; however, integration this time suggested an average of 48 repeats/glycan. An independent estimate of 47 LacNAc repeats was obtained by dynamic light scattering, which predicted an average molecular mass of 18.1 kDa.

Thus, taking into account the LacNAc repeat estimates of 59 (GC-MS monosaccharide composition), 54 (GC-MS methylation linkage), 48 (NMR resonance integration), and 47 (dynamic light scattering), we suggest an average figure of around 54 LacNAc repeats/oligosaccharide in the VLPL fraction. Unfortunately, but unsurprisingly, molecular species of this size (*i.e.* around 20.5 kDa) were not identified when VLPL was analyzed by ES-MS or MALDI-MS either before or after permethylation. On the other hand, smaller resolved glycans (*e.g.* the numbered peaks in Fig. 3*A*) were successfully analyzed by ES-MS (Table III). For exam-

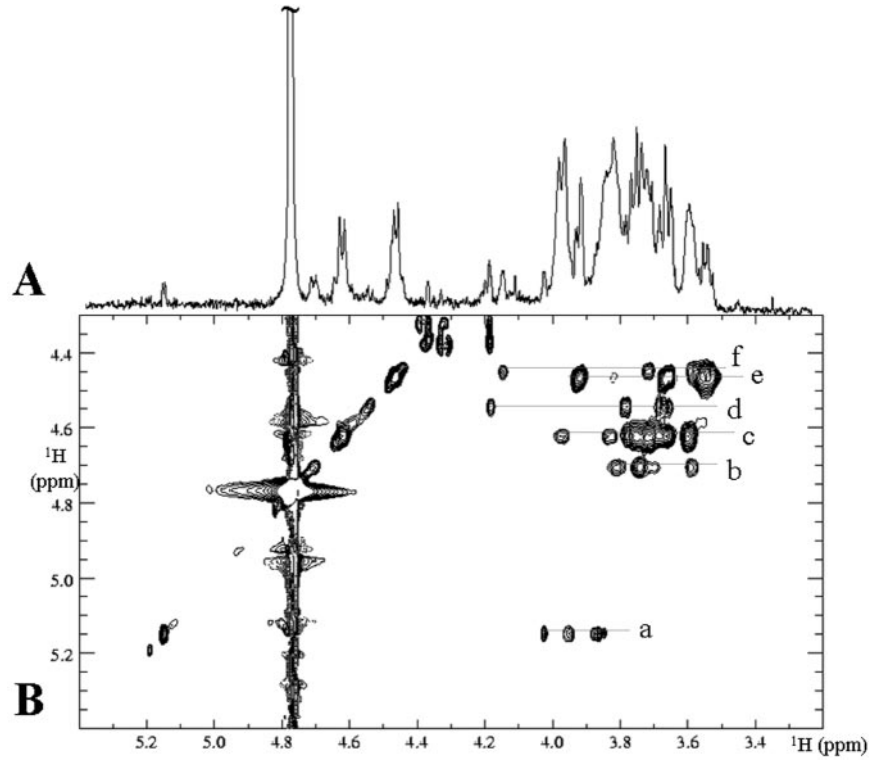


FIG. 4. **$^1$H NMR spectra of the VLPL fraction.** *A*, one-dimensional $^1$H NMR spectrum of the VLPL fraction (sample 1) after exchange into $^2H_2O$. *B*, anomeric region of the two-dimensional $^1$H TOCSY spectrum of the same sample. The peaks are *labeled* as in Table II.

TABLE II
*$^1$H NMR chemical shift assignments of the VLPL fraction and integrals of the anomeric protons*

VLPL fractions were exchanged into $^2H_2O$ by rotary evaporation (sample 1) and freeze-drying (sample 2) and analyzed by one-dimensional and two-dimensional $^1$H NMR, as described under "Materials and Methods."

| Label[a] | Residue[b] | $H_1$ | $H_2$ | $H_3$ | $H_4$ | $H_5$ | $H_{6,6'}$ | NAc | $H_1$ integral | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Sample 1 | Sample 2 |
| a | *$\alpha$Man | 5.15 | 3.87 | 3.96 | 4.02 | | | | 2 | 6.6[c] |
| b | *$\beta$GlcNAc | 4.71/4.70 | 3.82/3.79 | 3.59 | 3.71 | 3.74 | 3.83,3.95 | 1.91 | 2 | 2 |
| c | $\beta$GlcNAc | 4.63 | 3.77 | 3.60 | 3.72 | 3.67 | 3.83,3.97 | 1.91 | 22 | 48 |
| d | $\beta$Gal (3-substituted) | 4.55 | 3.66 | 3.79 | 4.18 | 3.55 | 3.68 | | 1.1 | 4.4 |
| e | $\beta$Gal (t- and 6-substituted) | 4.47/4.48 | 3.55 | 3.67 | 3.93 | | | | 25 | NM[d] |
| f | $\beta$Gal (3,6-di-substituted) | 4.45 | 3.58 | 3.72 | 4.15 | 3.57 | 3.91 | | | NM[d] |
| g | *$\beta$Man | ND[e] | 4.18 | 3.75 | | | | | | |

[a] Labels correspond with those shown in Fig. 4*B*.
[b] $Man_3GlcNAc_2$ core residues are marked with an asterisk.
[c] Integral affected by overlap with tail of water peak.
[d] Integral could not be measured accurately because of overlap with an impurity.
[e] Not detected because of overlap with the water resonance.

ple, an aliquot of peak 5 gave triply charged $[M+3Na]^{3+}$ ions at *m*/*z* 1544, 1665, and 1787. These ions correspond to $(LacNAc)_{10}Man_3GlcNAc_2$, $(LacNAc)_{11}Man_3GlcNAc_2$, and $(LacNAc)_{12}Man_3GlcNAc_2$, respectively (Table III). Daughter ion spectra of these ions showed intense *m*/*z* 366 daughter ions, consistent with poly-LacNAc structures.

The methylation linkage analysis was carefully quantified by determining molar relative response factors for the relevant partially methylated alditol acetate derivatives using a set of authentic standards. Thus, the data in (Table I) show that there is on average one -4GlcNAc$\beta$1–6(-4GlcNAc$\beta$1–3)Gal$\beta$1- branch point every six LacNAc repeats (*i.e.* 9 of 54 Gal residues are 3,6-disubstituted Gal residues) and that there is a 10:1 ratio of -4GlcNAc$\beta$1–6Gal$\beta$1- to -4GlcNAc$\beta$1–3Gal$\beta$1- interrepeat linkages.

*Tomato Lectin Chromatography*—Because tomato lectin is known to bind material from the flagellar pocket, it was suggested that glycoproteins from the flagellar pocket might contain linear poly-LacNAc chains (19). We examined which subset of the ricin-binding total poly-LacNAc fraction eluted from the void volume of Bio-Gel P-4 binds to tomato lectin. GC-MS monosaccharide analysis revealed that 85% of the applied *N*-glycans were bound by the tomato lectin column and subsequently eluted with a mixture of tri-*N*-acetylchitotriose and tetra-*N*-acetylchitotetraose. A selective loss of Man content was noted, suggesting that some of the smaller structures might not be retained by tomato lectin. The Dionex HPAEC chromatogram of the whole poly-LacNAc fraction (Fig. 3*A*) was compared with that of the tomato lectin-bound *N*-glycans (Fig. 3*B*). There is an obvious disappearance of glycans with retention time lower than 15 min. Peak 4 (Fig. 3, *A* and *B*, Table III), which contains sialic acid and is most likely derived from fetal calf serum glycoprotein, was also lacking in the tomato lectin Dionex HPAEC profile. Glycans typical of peaks 1, 2, and 4 (Table III and Fig. 3*B*) were detected in both the flow-through and the wash fraction by ES-MS (data not shown). Methylation linkage analysis of the tomato lectin-bound fraction revealed the presence of the same carbohydrate species present in the total fraction.

*Subcellular Localization of Ricin-binding Glycoproteins*—Sites of ricin binding were detected by fluorescence microscopy (Fig. 5*A*). Ricin binds the flagellar pocket, lysosomal/endosomal

TABLE III
*ES-MS analysis of HPAEC-resolved peaks from the total poly-LacNAc fraction*

Individual peak fractions from the HPAEC separation of the total poly-LacNAc fraction (see Fig. 3*A*) were adjusted to 50% acetonitrile, 1% formic acid and analyzed by nanospray positive ion ES-MS using a Q-Tof2 mass spectrometer. The proposed structures are consistent with the measured *m/z* values of the recorded double- and triple-charged ions.

| Peak[a] | ES-MS analysis | Proposed structure |
|---|---|---|
| 1 | $[M+2H]^{2+}$ = 821.30 | $LacNAc_2Man_3GlcNAc_2$ |
| 2 | $[M+2H]^{2+}$ = 1186.43 | $LacNAc_4Man_3GlcNAc_2$ |
| 3 | $[M+3Na]^{3+}$ = 1056.71 | $LacNAc_6Man_3GlcNAc_2$ |
| 4[b] | $[M+2H]^{2+}$ = 966.84 | $NeuAc_1LacNAc_2Man_3GlcNAc_2$ |
| 5 | $[M+3Na]^{3+}$ = 1543.55 | $LacNAc_{10}Man_3GlcNAc_2$ |
| | $[M+3Na]^{3+}$ = 1665.26 | $LacNAc_{11}Man_3GlcNAc_2$ |
| | $[M+3Na]^{3+}$ = 1786.97 | $LacNAc_{12}Man_3GlcNAc_2$ |

[a] Peak number refers to those shown in Fig. 3*A*.
[b] This minor sialic acid-containing structure is thought to originate from fetal calf serum glycoproteins.

system, and, more weakly, the general cell body. Binding of ricin is completely inhibited in the presence of the competing mono- and disaccharides Gal and lactose (Fig. 5*B*).

## DISCUSSION

The flagellar pocket of bloodstream form *T. brucei* is the only known site for endocytosis and secretion (4–6). Considerable demand is placed on the efficiency of the flagellar pocket to acquire nutrients, like serum transferrin, required by the cell to survive and propagate in the mammalian host. The flagellar pocket of *T. brucei* contains stage-specific glycoconjugates that bind ricin (18) and tomato lectin (19). In this work we exploited ricin affinity to purify a glycoprotein fraction that includes these molecules. Lectin fluorescence microscopy of fixed cells revealed that ricin-binding molecules are not restricted to the flagellar pocket but are also prevalent throughout the endocytic compartments and the lysosomes. A detailed proteomic analysis of the ricin-binding glycoproteins[2] revealed the presence of lysosomal/endosomal p67, flagellar pocket acid phosphatase, traces of VSG, and several other known and unknown glycoproteins. Thus, although the fraction we prepared clearly contains flagellar pocket matrix material, it is not representative of this material alone. Nevertheless, analysis of this ricin-binding material has provided several new insights into the glycobiology of bloodstream form *T. brucei*, as described below.

The approach of solubilizing cell ghosts with 8 M urea and 3% SDS allowed the complete extraction of high molecular mass glycoproteins that were not, in our hands, well extracted by detergents alone. This may explain why glycoproteins extracted in 1% CHAPS and purified by tomato lectin appear to have a more restricted molecular mass range (19). However, as reported for tomato lectin (19), the majority of the ricin-binding glycans of bloodstream form *T. brucei* glycoproteins are *N*-linked to asparagine residues. Some of the ricin-binding glycoproteins may form higher molecular mass complexes maintained by disulfide linkage, as judged by the effect of dithiothreitol reduction, but disulfide linkages do not appear to be extensive.

The VLPL poly-LacNAc *N*-linked glycans purified from bloodstream form *T. brucei* show several unusual features. First, their size is unique. With a mean LacNAc repeat number of about 54/glycan they are, to our knowledge, the largest poly-LacNAc structures yet reported. Previous reports for mammalian poly-LacNAc chains suggest a maximum of about 26 LacNAc repeats (22–24), and even *N*-linked keratan sulfate chains of articular cartilage appear to be limited to about 8 sulfated LacNAc repeats/antenna (31). The extreme size of the *T. brucei* poly-LacNAc glycans may explain our difficulties in redissolving them after rotary evaporation and suggests that the glycoproteins that carry these "giant" glycans may produce a gel-like matrix in the lumen of the flagellar pocket and/or the lumen of the endosomal/lysosomal system.

Second, the *T. brucei* poly-LacNAc *N*-linked glycans exhibit a high -4GlcNAcβ1-*6*Galβ1-:-4GlcNAcβ1-*3*Galβ1- inter-LacNAc repeat linkage ratio (~10:1). To our knowledge, there are no reports of this structural arrangement in any other eukaryote. For example, this ratio is reversed in mammalian poly-LacNAc structures (24, 32). The presence of 6-*O*-substituted Gal in a

[2] A. Atrih and M. A. J. Ferguson, unpublished data.

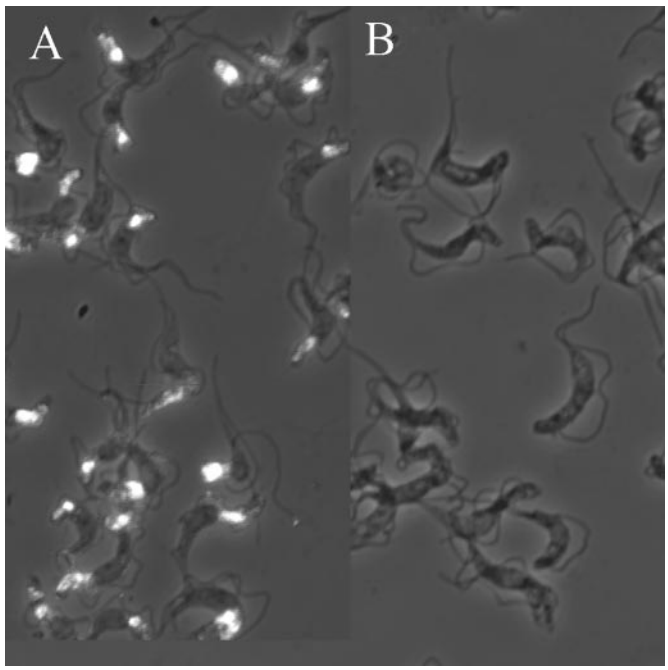


FIG. 5. **Subcellular localization of ricin-binding glycoproteins in bloodstream form *T. brucei* by fluorescence microscopy.** Fixed trypanosomes were stained with FITC-ricin (*A*) and FITC-ricin in the presence of the ricin-blocking sugars galactose and lactose (*B*). The fluorescence image is merged with a phase-contrast image.

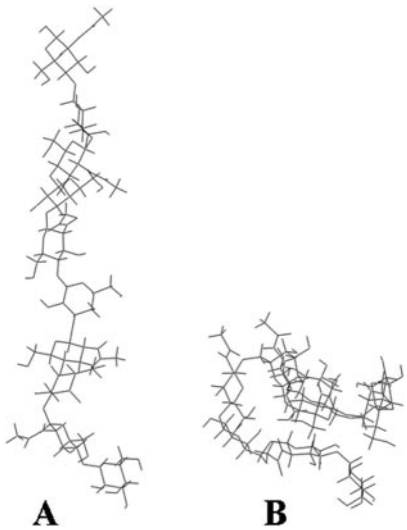


FIG. 6. **Molecular models of poly-LacNAc chains.** *A*, the extended conformation of five repeats of -4GlcNAcβ1-3Galβ1-linked poly-LacNAc (*i.e.* Galβ1–4GlcNAcβ1-(3Galβ1–4GlcNAcβ1-)$_3$3Galβ1–4GlcNAc) and (*B*) the more compact structure of five repeats of -4GlcNAcβ1-6Galβ1-linked poly-LacNAc (*i.e.* Galβ1–4GlcNAcβ1-(6Galβ1–4GlcNAcβ1-)$_3$6Galβ1–4GlcNAc).

linear poly-LacNAc sequence makes them resistant to digestion with endo-β-galactosidase (33). This property has prevented us from degrading the *T. brucei* structures into component fractions. We undertook molecular modeling of the low energy conformers of -4GlcNAcβ1-3Galβ1- and -4GlcNAcβ1-6Galβ1-linked poly-LacNAc chains. We found that the former adopted a generally extended conformation whereas the latter adopted a more compact structure (Fig. 6). This is consistent with the observation that structures containing -4GlcNAcβ1-3Galβ1- linkages have a higher hydrodynamic volume than those containing -4GlcNAcβ1-6Galβ1- linkages (25). This distinct and compact shape may play a significant role in the physical properties of the VLPL glycans and the glycoproteins to which they are attached. The dominance of -4GlcNAcβ1-6Galβ1-linked poly-LacNAc suggests that the enzymatic machinery involved in biosynthesis of poly-LacNAc in *T. brucei* differs from that of the mammalian systems. Indeed, BLAST searches in the *T. brucei* data base for homologs of the mammalian enzymes that initiate and participate in poly-LacNAc synthesis did not return obvious candidates. Thus, the poly-LacNAc biosynthetic machinery in *T. brucei* may represent a potential therapeutic target. It is conceivable that the essentiality of galactose metabolism to bloodstream form *T. brucei* (34) may be linked to an absolute requirement for the synthesis of giant poly-LacNAc chains by the parasite.

Third, despite their size, the *T. brucei* poly-LacNAc *N*-linked glycans are compositionally very simple. They are neutral (nonsialylated, nonsulfated) molecules that contain only Man, Gal, and GlcNAc; fucose, a common component of complex *N*-linked glycans, is noticeably absent.

Fourth, the poly-LacNAc structures of the VLPL fraction are distinct from the smaller (≤ 8 LacNAc repeats) *N*-linked glycans found attached to certain type-2 VSGs (25). The latter contain mostly branched poly-LacNAc chains (with some linear but predominantly -4GlcNAcβ1-3Galβ1-linked repeats) and can, like the complex biantennary glycans of type-3 VSGs, terminate in Galα1–3Galβ1- (25). Interestingly, despite an average of 9 branch points/VLPL glycan, which predicts an average of 11 nonreducing β-galactose termini/glycan, there is no evidence for Galα1–3Galβ1- termini in these structures. This may be immunologically relevant because normal human serum contains a substantial amount of antibody directed against the Galα1–3Galβ1- epitope (35). Thus, whereas the Galα1–3Galβ1-containing VSG glycans are buried in the VSG coat and cannot be reached by circulating antibodies (36), at least some of the VLPL fraction is exposed to host serum in the flagellar pocket. Another possible reason for differences in the VLPL and VSG poly-LacNAc structures may be to do with intracellular sorting. Nolan *et al.* (19) have postulated that poly-LacNAc chains in bloodstream form *T. brucei* may play a role as sorting signals in the endocytotic pathway. It is possible that the size and/or the ratio of 6-*O*-substituted Gal to 3-*O*-substituted Gal in poly-LacNAc glycans may be a crucial factor for recognition by putative lectin-like receptor(s). This might allow these receptors to differentiate between their true cargo and the ubiquitous VSG that coats the entire flagellar pocket.

Lastly, although tomato lectin is known to bind with high affinity to glycopeptides containing three or more linear LacNAc repeats linked via -4GlcNAcβ1-3Galβ1- inter-LacNAc linkages (32, 37), our results suggest that structures with 3 or more linear LacNAc repeats with -4GlcNAcβ1-6Galβ1- inter-LacNAc linkages are also bound by the lectin. Previously, such structures were not available to test this specificity.

*Acknowledgments*—We thank Irene Hallyburton (Post-Genomics and Molecular Interactions Centre) for performing the light scattering studies and Angela Mehlert and Lucia Güther for advice and assistance.

REFERENCES

1. Mehlert, A., Zitzmann, N., Richardson, J. M., Treumann, A., and Ferguson, M. A. (1998) *Mol. Biochem. Parasitol.* **91,** 145–152
2. Cross, G. A. M. (1996) *BioEssays* **18,** 283–291
3. Pays, E., and Nolan, D. P. (1998) *Mol. Biochem. Parasitol.* **91,** 3–36
4. Lee, M. G. S., Russell, D. G., D'Alesandro, P. A., and Van der Ploeg, L. X. T. (1994) *J. Biol. Chem.* **269,** 8408–8415
5. Overath, P., Stierhof, Y. D., and Wiese, M. (1997) *Trends Cell Biol.* **7,** 27–33
6. Landfear, S. M., and Ignatushchenko, M. (2001) *Mol. Biochem. Parasitol.* **115,** 1–17
7. McConville, M. J., Mullin, K. A., Ilgoutz, S. C., and Teasdale, R. D. (2002) *Microbiol. Mol. Biol. Rev.* **66,** 122–154
8. Geuskens, M., Pays, E., and Cardoso de Almeida, M. L. (2000) *Mol. Biochem. Parasitol.* **108,** 269–275
9. Jeffries, T. R., Morgan, G. W., and Field, M. C. (2000) *J. Cell Sci.* **114,** 2617–2626
10. Webster, P., and Russell. D. (1993) *Parasitol. Today* **9,** 201–205
11. Engstler, M., Thilo, L., Weise, F., Grunfelder, C. G., Schwarz, H., Boshart, M., and Overath, P. (2004) *J. Cell Sci.* **117,** 1105–1115
12. Webster, P., and Fish, W. R. (1989) *Eur. J. Cell Biol.* **49,** 303–310
13. Morgan, G. W., Allen, C. L., Jeffries, T. R., Hollinshead, M., and Field, M. C. (2001) *J. Cell Sci.* **114,** 2605–2615
14. Morgan, G. W., Hall, B. S., Denny, P. W., Carrington, M., and Field, M. C. (2002) *Trends Parasitol.* **18,** 491–496
15. Liu, J., Qiao, X., Du, D., and Lee, M. G. (2000) *J. Biol. Chem.* **275,** 12032–12040
16. Alexander, D. L., Schartz, K. J., Balber, A. E., and Bangs, J. D. (2002) *J. Cell Sci.* **115,** 3253–3263
17. Balber, A. E. (1990) *Crit. Rev. Immunol.* **10,** 177–201
18. Brickman, M. J., and Baber, A. E. (1990) *J. Protozool.* **37,** 219–224
19. Nolan, D. P., Geuskens, M., and Pays, E. (1999) *Curr. Biol.* **9,** 1169–1172
20. Nolan, D. P., Jackson, D. G., Windle, H. J., Pays, A., Geuskens, M., Michel, A., Voorheis, H. P., and Pays, E. (1997) *J. Biol. Chem.* **272,** 29212–29221
21. Schell, D., Stierhof, Y. D., and Overath, P. (1990) *FEBS Lett.* **271,** 67–70
22. Carlsson, S. R., and Fukuda, M. (1990) *J. Biol. Chem.* **265,** 20488–20495
23. Helenius, A., and Aebi, M. (2001) *Science* **291,** 2364–2369
24. Fukuda, M. N., Dell, A., Oates, J. E., and Fukuda, M. (1985) *J. Biol. Chem.* **260,** 6623–6631
25. Zamze, S. E., Ashford, D. A., Wooten, E. W., Rademacher, T. W., and Dwek, R. A. (1991) *J. Biol. Chem.* **266,** 20244–20261
26. Cross, G. A. M. (1984) *J. Cell. Biochem.* **24,** 79–90
27. Ferguson, M. A. J. (1992) in *Lipid Modification of Proteins: A Practical Approach* (Turner, A. J., and Hooper, N., eds) pp 191–203, IRL Press, Oxford
28. Ferguson, M. A. J., Homans, S. W., Dwek, R. A., and Rademacher, T. W. (1988) *Science* **239,** 753–759
29. Mehlert, A., Richardson, J. M., and Ferguson, M. A. J. (1998) *J. Mol. Biol.* **277,** 379–392
30. Geyer, A., Fitzpatrick, T. B., Pawelek, P. D., Kitzing, K., Vrielink, A., Ghisla, S., and Macheroux, P. (2001) *Eur. J. Biochem.* **268,** 4044–4053
31. Lauder, R. M., Huckerby, T. N., Brown, G. M., Bayliss, M. T., and Nieduszynski, I. A. (2001) *Biochem. J.* **358,** 523–528
32. Lee, N., Wang, W.-C., and Fukuda, M. (1990) *J. Biol. Chem.* **265,** 20476–20487
33. Renkonen, O., Penttila, L., Makkonen, A., Niemela, R., Leppanen, A., Helin, J., and Vainio, A. (1989) *Glycoconj. J.* **6,** 129–140
34. Roper, J. R., Güther, M. L. S., Milne, K. G., and Ferguson, M. A. J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5884–5889
35. Galili, U., Rachmilewitz, E. A., Peleg, A., and I. Flechner, I. (1984) *J. Exp. Med.* **160,** 1519–1531
36. Mehlert, A., Bond, C. S., and Ferguson, M. A. (2002) *Glycobiology* **12,** 607–612
37. Merkle, R. K., and Cummings, R. D. (1987) *J. Biol. Chem.* **262,** 8179–8187